International Journal of Bio-Technology and Research (IJBTR)
ISSN (P):2249-6858; ISSN (E):2249-796X
Vol. 9, Issue 1, Jun 2019, 5-8

COMPARATIVE GENOMICS - A PERSPECTIVE

ESHA DOGRA & PRASHANT SINGH

Department of Biotechnology, Lovely Professional University, Phagwara, Punjab, India

ABSTRACT

The rapidly emerging field of comparative genomics has yielded dramatic results. Comparative genome analysis has become feasible with the availability of a number of completely sequenced genomes. Comparison of complete genomes between organisms allow for global views on genome evolution and the availability of many completely sequenced genomes increases the predictive power in deciphering the hidden information in genome design, function and evolution. Thus, comparison of human genes with genes from other genomes in a genomic landscape could help assign novel functions for un-annotated genes. The MicroRNAs are sequence-specific regulators of post-transcriptional gene expression in many eukaryotes. They are believed to control the expression of thousands of target mRNAs, with each mRNA believed to be targeted by multiple MicroRNAs. In this study it has been demonstrated about the construction of a cut off for sequence similarity and query coverage on the basis of known datasets of Pan troglodytes (chimpanzee) miRNA by performing genomic nucleotide BLAST with Homo sapiens (human) genome. The two combined approaches help in the prediction of 909 miRNA candidates in chimpanzee similar to that of humans.

KEYWORDS: Genome, MicroRNAs, Pan Troglodytes & Homo Sapiens

Received: Dec 22, 2018; **Accepted:** Jan 13, 2019; **Published:** Jan 24, 2019; **Paper Id.:** IJBTRJUN20192

Original Article

INTRODUCTION

RNA is a nucleic acid polymer which consists of nucleotide as monomers which allow RNA to encode genetic information. RNA is mostly single stranded molecule which folds to form secondary structure RNA. According to the central dogma of molecular biology RNA acts as a messenger between DNAand the protein synthesis. However, this central dogma is getting challenged by the recent finding that tiny fragments i.e. the non coding RNA microRNA (miRNA) of 19-25 nucleotides in length present in the genomes of plants and animals are able to negatively regulate protein-coding genes by interfering with mRNA's original instructions Zhang et al. (2005).miRNAs are the highly conserved regions which suppresses the gene expression by imperfect base pairing to the 3´untranslated region (UTR) oftarget mRNAs which leads to the repression of protein production or mRNA degradation Bentwich et al. (2005).

MicroRNAs are related to, but distinct from, short interfering RNAs (siRNAs). A key difference between siRNA and microRNA is that siRNA requires almost complete complementary to its targeting sequence for it to exert the silencing function, whereas a microRNA usually binds to its target genes through partial complementary binding. Because of this unique feature, a single microRNA has multiple target genes and, thus, could regulate a large number of protein-coding genes. This may explain why microRNAs play a fundamental role in the regulation of diverse cellular processes. Increasing efforts to identify specific targets of microRNAs have led to the speculation that microRNAs may regulate at least 30% of human protein encoding genes. The first miRNA, lin-4, was identified by Lee et al.(1993) in a genetic screen for mutants that disrupt the timing of post-embryonic

development in *Caenorhabditiselegans* Lee RC and Ambros V (2001). The rapid progress in genome sequencing demands more comparative analysis to gain new insights into evolutionary, biochemical, genetic, metabolic, and physiological pathways. Comparative genomics is the direct comparison of complete genetic material of one organism against that of another to gain a better understanding of how species evolved and to determine the function of genes and noncoding regions in genomes. It includes a comparison of gene number, gene content, and gene location, the length and number of coding regions (called exons) within genes, the amount of non coding DNA in each genome, and conserved regions maintained in both prokaryotic and eukaryotic groups of organisms. Comparative genomics not only can trace out the evolutionary relationship between organisms, but also differences and similarities within and between species.

MicroRNAs (miRNAs) are small non-coding RNAs, 19–24 nt long, which play a crucial regulatory role by inhibiting the translation of protein-coding mRNAs in various eukaryotic organisms. A miRNAs is processed from a longer transcript, referred to as the primary transcript (pri-miRNA). miRNAs can be located within the introns of protein-coding genes, outside of protein-coding genes entirely ('intergenic') or more rarely in a coding exon, untranslated region (UTR) or exon of a non-coding transcript. These tiny miRNAs inhibit the translation of a mRNA into protein through imperfect base pairing to one or more target sequences in the mRNA. The identification of animal miRNA targets is a challenging assignment for both experimental and computational groups.

MATERIALS AND METHODS

Formulation of Hypotheses

- Identification of the miRNAs of Pan troglodytes using sequence similarity approach by NCBI eukaryotic genomic BLAST with respect to the miRNAs of Homo sapiens.
- Formation of threshold for the prediction of miRNAs of Pan troglodytes after sequence similarity from already known miRNAs of Pan troglodytes.
- Identification of the miRNAs by considering the flanking regions of the miRNAs having less similarity than that of the cut off.

METHODOLOGY

There are 1048 miRNAs of human present in miRBase database. So to find out the sequence similarities between human and chimpanzee various steps were followed. A **Collection of all the 1048 pre-miRNAs sequence of *Homo sapiens* from the miRBase web server, Perform genomic BLASTn, threshold formation, flanking region study**

RESULTS AND DISCUSSION

Sequence Similarity Results

All the 1048 pre-miRNA fasta sequences of *Homo sapiens* were collected from miRBase web server and saved in a Microsoft word file. With each sequence NCBI genomic BLASTn was performed with respect to the genome of *Pan troglodytes* and saved the results in Microsoft excel sheet and analyzed the results by data filtering (Table 1 and 2).

Threshold Formation

The 601 pre-miRNA fasta sequences of *Pan troglodytes* (chimpanzee) were collected from miRBase web server and saved in a Microsoft word file. With each sequence genomic BLASTn was performed with respect to the genome of

Homo sapiens (human)and saved the results in Microsoft excel sheet and analysed the results by data filtering (Table 3).

Flanking Region Study

The flanking regions (left as well as right) of *Homo sapiens* and *Pan troglodytes* were identified for the sequences having <95% maximum identity analysed in the first step. The length of the flanking sequence was taken to be 100bps nucleotide. It is known that the pre-miRNA sequence is 60-80bps in length; therefore the length of the flanking sequences for the flanking region similarity study was taken as 100bps nucleotides. The BLAST was then performed between the flanking sequence of human and chimpanzee (left as well as right) and the results were saved in a Microsoft excel sheet and analysed by data filtering (Table 4)

CONCLUSIONS

The comparative genomics approach analysis helps us in the prediction of pre-miRNA in case of *Pan troglodytes.* Till now there are only 601 miRNAs in *Pan troglodytes* which are predicted according to the most reliable web server i.e. miRBase database. There are 1048 miRNAs present in *Homo sapiens*. There are possibilities of having more miRNAs in chimpanzee as it is the closest species to that of humans. After performing sequence similarity of Human miRNA with that of chimpanzee genome 887 pre-miRNAs had been predicted which are similar to human miRNAs. Similarly, flanking region study was done to predict further miRNA similarity and 22 more pre-miRNAs are identified. Therefore, 909 pre-miRNAs are predicted in chimpanzee. miRNA dys functioning is related to various diseases. So studying these miRNAs can help us to find proper treatment and valuable medication for trials. miRNA has been mostly conserved over a varied number of species. Over 99% of MiRNA are conserved. This prediction will lead us to develop new thoughts for further prediction of unknown functions of miRNA. These miRNAs will further help in the evolutionary relationship among different organisms. miRNA research has been conducted for only a few years. There is still lots of unknown but exciting knowledge to be revealed about miRNAs.

REFERENCES

1. *Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, Barzilai A, Einat P, Einav U and Meiri E (2005). Identification of hundreds of conserved and nonconserved human microRNAs. Nat. Genet 37: 766–770.*
2. *Lee RC and Ambros V (2001). An extensive class of small RNAs in Caenorhabditis elegans. Scien. 294: 862–864*
3. *Lee RC, Feinbaum RL and Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14.Cell 75: 843-854.*
4. *Kumari, U., & Choudhary, A. K. (2016). Genome Sequence Analysis of Solanum Lycopersicum by Applying Sequence Alignment Method to Determine the Statistical Significance of an Alignment.*
5. *Zhang LV, King OD, Wong SL, Goldberg DS, Tong AH, Lesage G, Andrews B, Bussey H, Boone C, Roth FP. 2005. Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network. J Biol 4: 6.*

APPENDIX

Table 1: Data Obtained by Sequence Similarity Study of *Homo sapiens* with Respect to *Pan troglodyte* (Chimpanzee) Genome

Maximum Identity	No. of miRNAs
91%-100%	949
81%-90%	10
<=80%	0

Table 2: Data Obtained by Sequence Similarity Study of *Homo sapiens* with Respect to *Pan troglodyte* (Chimpanzee) Genome Considering Query Coverage

Maximum Identity	Query Coverage	No. of miRNAs
100%	100%	722
98%-99%	100%	2
95%-97%	100%	116
90%-99%	90%-99%	40
<90%	<90%	3
0%		89

Table 3: Data obtained by Sequence Similarity Study of *Pan troglodytes* with that of *Homo sapiens* Genome

Maximum Identity	Query Coverage	No. of miRNAs
100%	100%	375
98%-99%	100%	125
95%-97%	100%	52
91%-94%	100%	2
85%-90%	100%	1
<85%		0
0%		33

Table 4: Data Obtained for the Flanking Sequence Similarity

Flanking Region	Maximum Identity >80% and Query Coverage >90%
Both left and right	22